Oxford Handbooks Online

Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques

John Gerring

The Oxford Handbook of Political Methodology
Edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier

Print Publication Date: Aug 2008
Subject: Political Science, Comparative Politics, Political Methodology
Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780199286546.003.0028

Abstract and Keywords

This article presents some guidance by cataloging nine different techniques for case selection: typical, diverse, extreme, deviant, influential, crucial, pathway, most similar, and most different. It also indicates that if the researcher is starting from a quantitative database, then methods for finding influential outliers can be used. In particular, the article clarifies the general principles that might guide the process of case selection in case-study research. Cases are *more or less* representative of some broader phenomenon and, on that score, may be considered better or worse subjects for intensive analysis. The article then draws attention to two ambiguities in case-selection strategies in case-study research. The first concerns the admixture of several case-selection strategies. The second concerns the changing status of a case as a study proceeds. Some case studies follow only one strategy of case selection.

Keywords: typical case, diverse case, extreme case, deviant case, influential case, crucial case, pathway case, most similar case, most different case, case-study analysis

CASE-study analysis focuses on one or several cases that are expected to provide insight into a larger population. This presents the researcher with a formidable problem of case selection: Which cases should she or he choose?

In large-sample research, the task of case selection is usually handled by some version of randomization. However, in case-study research the sample is small (by definition) and this makes random sampling problematic, for any given sample may be wildly unrepresentative. Moreover, there is no guarantee that a few cases, chosen randomly, will provide leverage into the research question of interest.

In order to isolate a sample of cases that both reproduces the relevant causal features of a larger universe (representativeness) and provides variation along the dimensions of theoretical interest (causal leverage), case selection for very small samples must employ

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

purposive (nonrandom) selection procedures. Nine such methods are discussed in this chapter, each of which may be identified with a distinct case-study (p. 646) "type:" *typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar*, and *most-different*. Table 28.1 summarizes each type, including its general definition, a technique for locating it within a population of potential cases, its uses, and its probable representativeness.

While each of these techniques is normally practiced on one or several cases (the diverse, most-similar, and most-different methods require at least two), all *may* employ additional cases—with the proviso that, at some point, they will no longer offer an opportunity for in-depth analysis and will thus no longer be "case studies" in the usual sense (Gerring 2007, ch. 2). It will also be seen that small-*N* case-selection procedures rest, at least implicitly, upon an analysis of a larger population of potential cases (as does randomization). The case(s) identified for intensive study is chosen from a population and the reasons for this choice hinge upon the way in which it is situated within that population. This is the origin of the terminology—typical, diverse, extreme, et al. It follows that case-selection procedures in case-study research may build upon prior cross-case analysis and that they depend, at the very least, upon certain assumptions about the broader population.

In certain circumstances, the case-selection procedure may be structured by a quantitative analysis of the larger population. Here, several caveats must be satisfied. First, the inference must pertain to more than a few dozen cases; otherwise, statistical analysis is problematic. Second, relevant data must be available for that population, or a significant sample of that population, on key variables, and the researcher must feel reasonably confident in the accuracy and conceptual validity of these variables. Third, all the standard assumptions of statistical research (e.g. identification, specification, robustness) must be carefully considered, and wherever possible, tested. I shall not dilate further on these familiar issues except to warn the researcher against the unreflective use of statistical techniques.[1] When these requirements are *not* met, the researcher must employ a qualitative approach to case selection.

The point of this chapter is to elucidate general principles that might guide the process of case selection in case-study research, building upon earlier work by Harry Eckstein, Arend Lijphart, and others. Sometimes, these principles can be applied in a quantitative framework and sometimes they are limited to a qualitative framework. In either case, the logic of case selection remains quite similar, whether practiced in small-*N* or large-*N* contexts.

Before we begin, a bit of notation is necessary. In this chapter "*N*" refers to cases, not observations. Here, I am concerned primarily with causal inference, rather than inferences that are descriptive or predictive in nature. Thus, all hypotheses involve at least one independent variable (X) and one dependent variable (Y). For convenience, I shall label the causal factor of special theoretical interest X_1, and the control variable, or vector of controls (if there are any), X_2. If the writer is concerned to explain a puzzling outcome, but has no preconceptions about its causes, then the research will be described

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

as *Y-centered*. If a researcher is concerned to (p. 647) (p. 648) investigate the effects of a particular cause, with no preconceptions about what these effects might be, the research will be described as *X-centered*. If a researcher is concerned to investigate a particular causal relationship, the research will be described as X_1/*Y-centered*, for it connects a particular cause with a particular outcome.[2] *X*- or *Y*-centered research is exploratory; its purpose is to generate new hypotheses. X_1/*Y-centered* research, by contrast, is confirmatory/disconfirmatory; its purpose is to test an existing hypothesis.

Table 28.1. Techniques of case selection

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

1. Typical

º *Definition*: Cases (1 or more) are typical examples of some cross-case relationship.

º *Cross-case technique*: A low-residual case (on-lier).

º *Uses*: Hypothesis-testing.

º *Representativeness*: By definition, the typical case is representative.

2. Diverse

º *Definition*: Cases (2 or more) illuminate the full range of variation on $X_{1,}$, Y, or X_1/Y.

º *Cross-case technique*: Diversity may be calculated by (a) categorical values of X_1 or Y (e.g. Jewish, Catholic, Protestant), (b) standard deviations of X_1 or Y (if continuous), (*c*) combinations of values (e.g. based on cross-tabulations, factor analysis, or discriminant analysis).

º *Uses*: Hypothesis generating or hypothesis testing.

º *Representativeness*: Diverse cases are likely to be representative in the minimal sense of representing the full variation of the population (though they might not mirror the *distribution* of that variation in the population).

3. Extreme

º *Definition*: Cases (1 or more) exemplify extreme or unusual values of X_1 or Y relative to some univariate distribution.

º *Cross-case technique*: A case lying many standard deviations away from the mean of X_1 or Y.

º *Uses*: Hypothesis-generating (open-ended probe of X_1 or Y).

º *Representativeness*: Achievable only in comparison with a larger sample of cases.

4. Deviant

º *Definition*: Cases (1 or more) deviate from some cross-case relationship.

º *Cross-case technique*: A high-residual case (outlier).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

º *Uses*: Hypothesis-generating (to develop new explanations for Y).

º *Representativeness*: After the case study is conducted it may be corroborated by a cross-case test, which includes a general hypothesis (a new variable) based on the case-study research. If the case is now an on-lier, it may be considered representative of the new relationship.

5. Influential

º *Definition*: Cases (1 or more) with influential configurations of the independent variables.

º *Cross-case technique*: Hat matrix or Cook's Distance.

º *Uses*: Hypothesis-testing (to verify the status of cases that may influence the results of a cross-case analysis).

º *Representativeness*: Not pertinent, given the goals of the influential-case study.

6. Crucial

º *Definition*: Cases (1 or more) are most or least likely to exhibit a given outcome.

º *Cross-case technique*: Qualitative assessment of relative crucialness.

º *Uses*: Hypothesis-testing (confirmatory or disconfirmatory).

º *Representativeness*: Often difficult to assess.

7. Pathway

º *Definition*: Cases (1 or more) that embody a distinct causal path from X_1 to Y.

º *Cross-case technique*: Cross-tab (for categorical variables) or residual analysis (for continuous variables).

º *Uses*: Hypothesis-testing (to probe causal mechanisms).

º *Representativeness*: May be tested by examining residuals for the chosen cases.

8. Most-similar

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

º *Definition /*: Cases (2 or more) are similar on specified variables other than X / $_{1}$ / and/or Y.

º *Cross-case technique*: Matching.

º *Uses*: Hypothesis-generating or hypothesis-testing.

º *Representativeness*: May be tested by examining residuals for the chosen cases.

9. Most-different

º *Definition*: Cases (2 or more) are different on specified variables other than X *1* and Y.

º *Cross-case technique*: The inverse of the most-similar method of large-N case selection (see above).

º *Uses*: Hypothesis-generating or hypothesis-testing (eliminating deterministic causes).

º *Representativeness*: May be tested by examining residuals for the chosen cases.

1 Typical Case

In order for a focused case study to provide insight into a broader phenomenon it must be representative of a broader set of cases. It is in this context that one may speak of a *typical-case* approach to case selection. The typical case exemplifies what is considered to be a typical set of values, given some general understanding of a phenomenon. By construction, the typical case is also a representative case.

Some typical cases serve an exploratory role. Here, the author chooses a case based upon a set of descriptive characteristics and then probes for causal relationships. Robert and Helen Lynd (1929/1956) selected a single city “to be as representative as possible of contemporary American life.” Specifically, they were looking for a city with

> 1) a temperate climate; 2) a sufficiently rapid rate of growth to ensure the presence of a plentiful assortment of the growing pains accompanying contemporary social change; 3) an industrial culture with modern, high-speed machine production; 4) the absence of dominance of the city's industry by a single plant (i.e., not a one-industry town); 5) a substantial local artistic (p. 649) life to balance its industrial activity …; and 6) the absence of any outstanding peculiarities or acute local problems which would mark the city off from the

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

> midchannel sort of American community. (Lynd and Lynd 1929/1956, quoted in Yin 2004, 29–30)

After examining a number of options the Lynds decided that Muncie, Indiana, was more representative than, or at least as representative as, other midsized cities in America, thus qualifying as a typical case.

This is an inductive approach to case selection. Note that typicality may be understood according to the mean, median, or mode on a particular dimension; there may be multiple dimensions (as in the foregoing example); and each may be differently weighted (some dimensions may be more important than others). Where the selection criteria are multidimensional and a large sample of potential cases is in play, some form of factor analysis may be useful in identifying the most-typical case(s).

However, the more common employment of the typical-case method involves a *causal* model of some phenomenon of theoretical interest. Here, the researcher has identified a particular outcome (Y), and perhaps a specific X_1/Y hypothesis, which she wishes to investigate. In order to do so, she looks for a typical example of that causal relationship. Intuitively, one imagines that a case selected according to the mean values of all parameters must be a typical case relative to some causal relationship. However, this is by no means assured.

Suppose that the Lynds were primarily interested in explaining feelings of trust/distrust among members of different social classes (one of the implicit research goals of the *Middletown* study). This outcome is likely to be affected by many factors, only some of which are included in their six selection criteria. So choosing cases with respect to a causal hypothesis involves, first of all, identifying the relevant parameters. It involves, secondly, the selection of a case that has a "typical" value relative to the overall causal model; it is well explained. Cases with untypical scores on a particular dimension (e.g. very high or very low) may still be typical examples of a causal relationship. Indeed, they may be more typical than cases whose values lie close to the mean. Thus, a descriptive understanding of typicality is quite different from a causal understanding of typicality. Since it is the latter version that is more common, I shall adopt this understanding of typicality in the remainder of the discussion.

From a qualitative perspective, causal typicality involves the selection of a case that conforms to expectations about some general causal relationship. It performs as expected. In a quantitative setting, this notion is measured by the size of a case's residual in a large-N cross-case model. Typical cases lie on or near the regression line; their residuals are small. Insofar as the model is correctly specified, the size of a case's residual (i.e. the number of standard deviations that separate the actual value from the fitted value) provides a helpful clue to how representative that case is likely to be. "Outliers" are unlikely to be representative of the target population.

Of course, just because a case has a low residual does not *necessarily* mean that it is a representative case (with respect to the causal relationship of interest). Indeed, **(p. 650)** the issue of case representativeness is an issue that can never be definitively settled.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

When one refers to a “typical case” one is saying, in effect, that the *probability* of a case's representativeness is high, relative to other cases. This test of typicality is misleading if the statistical model is mis-specified. And it provides little insurance against errors that are purely stochastic. A case may lie directly on the regression line but still be, in some important respect, atypical. For example, it might have an odd combination of values; the interaction of variables might be different from other cases; or additional causal mechanisms might be at work. For this reason, it is important to supplement a statistical analysis of cases with evidence drawn from the case in question (the case study itself) and with our deductive knowledge of the world. One should never judge a case solely by its residual. Yet, all other things being equal, a case with a low residual is less likely to be unusual than a case with a high residual, and to this extent the method of case selection outlined here may be a helpful guide to case-study researchers faced with a large number of potential cases.

By way of conclusion, it should be noted that because the typical case embodies a typical value on some set of causally relevant dimensions, the variance of interest to the researcher must lie *within* that case. Specifically, the typical case of some phenomenon may be helpful in exploring causal mechanisms and in solving identification problems (e.g. endogeneity between X_1 and Y, an omitted variable that may account for X_1 *and* Y, or some other spurious causal association). Depending upon the results of the case study, the author may confirm an existing hypothesis, disconfirm that hypothesis, or reframe it in a way that is consistent with the findings of the case study. These are the uses of the typical-case study.

2 Diverse Cases

A second case-selection strategy has as its primary objective the achievement of maximum variance along relevant dimensions. I refer to this as a *diverse-case* method. For obvious reasons, this method requires the selection of a set of cases—at minimum, two—which are intended to represent the full range of values characterizing X_1, Y, or some particular X_1/Y relationship.[3]

Where the individual variable of interest is categorical (on/off, red/black/blue, Jewish/Protestant/Catholic), the identification of diversity is readily apparent. The investigator simply chooses one case from each category. For a continuous variable, (p. 651) the choices are not so obvious. However, the researcher usually chooses both extreme values (high and low), and perhaps the mean or median as well. The researcher may also look for break-points in the distribution that seem to correspond to categorical differences among cases. Or she may follow a theoretical hunch about which threshold values count, i.e. which are likely to produce different values on Y.

Another sort of diverse case takes account of the values of multiple variables (i.e. a vector), rather than a single variable. If these variables are categorical, the identification of causal types rests upon the intersection of each category. Two dichotomous variables produce a matrix with four cells. Three trichotomous variables produce a matrix of eight

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

cells. And so forth. If all variables are deemed relevant to the analysis, the selection of diverse cases mandates the selection of one case drawn from within each cell. Let us say that an outcome is thought to be affected by sex, race (black/white), and marital status. Here, a diverse-case strategy of case selection would identify one case within each of these intersecting cells—a total of eight cases. Things become slightly more complicated when one or more of the factors is continuous, rather than categorical. Here, the diversity of case values do not fall neatly into cells. Rather, these cells must be created by fiat—e.g. high, medium, low.

It will be seen that where multiple variables are under consideration, the logic of diverse-case analysis rests upon the logic of *typological* theorizing—where different combinations of variables are assumed to have effects on an outcome that vary across types (Elman 2005; George and Bennett 2005, 235; Lazarsfeld and Barton 1951). George and Smoke, for example, wish to explore different types of deterrence failure—by "fait accompli," by "limited probe," and by "controlled pressure." Consequently, they wish to find cases that exemplify each type of causal mechanism.[4]

Diversity may thus refer to a range of variation on X or Y, or to a particular combination of causal factors (with or without a consideration of the outcome). In each instance, the goal of case selection is to capture the full range of variation along the dimension(s) of interest.

Since diversity can mean many things, its employment in a large-N setting is necessarily dependent upon how this key term is defined. If it is understood to pertain only to a single variable (X_1 or Y), then the task is fairly simple. A categorical variable mandates the choice of at least one case from each category—two if dichotomous, three if trichotomous, and so forth. A continuous variable suggests the choice of at least one "high" and "low" value, and perhaps one drawn from the mean or median. But other choices might also be justified, according to one's hunch about the underlying causal relationship or according to natural thresholds found in the data, which may be grouped into discrete categories. Single-variable traits are usually easy to discover in a large-N setting through descriptive statistics or through visual inspection of the data.

(p. 652) Where diversity refers to particular *combinations* of variables, the relevant cross-case technique is some version of stratified random sampling (in a probabilistic setting) or Qualitative Comparative Analysis (in a deterministic setting) (Ragin 2000). If the researcher suspects that a causal relationship is affected not only by combinations of factors but also by their *sequencing*, then the technique of analysis must incorporate temporal elements (Abbott 2001; Abbott and Forrest 1986; Abbott and Tsay 2000). Thus, the method of identifying causal types rests upon whatever method of identifying causal relationships is employed in the large-N sample.

Note that the identification of distinct case types is intended to identify groups of cases that are internally homogeneous (in all respects that might affect the causal relationship of interest). Thus, the choice of cases within each group should not be problematic, and may be accomplished through random sampling or purposive case selection. However, if

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

there is suspected diversity within each category, then measures should be taken to assure that the chosen cases are typical of each category. A case study should not focus on an atypical member of a subgroup.

Indeed, considerations of diversity and typicality often go together. Thus, in a study of globalization and social welfare systems, Duane Swank (2002) first identifies three distinctive groups of welfare states: "universalistic" (social democratic), "corporatist conservative," and "liberal." Next, he looks within each group to find the most-typical cases. He decides that the Nordic countries are more typical of the universalistic model than the Netherlands since the latter has "some characteristics of the occupationally based program structure and a political context of Christian Democratic-led governments typical of the corporatist conservative nations" (Swank 2002, 11; see also Esping-Andersen 1990). Thus, the Nordic countries are chosen as representative cases within the universalistic case type, and are accompanied in the case-study portion of his analysis by other cases chosen to represent the other welfare state types (corporatist conservative and liberal).

Evidently, when a sample encompasses a full range of variation on relevant parameters one is likely to enhance the representativeness of that sample (relative to some population). This is a distinct advantage. Of course, the inclusion of a full range of variation may distort the actual distribution of cases across this spectrum. If there are more "high" cases than "low" cases in a population and the researcher chooses only one high case and one low case, the resulting sample of two is not perfectly representative. Even so, the diverse-case method probably has stronger claims to representativeness than any other small-*N* sample (including the standalone typical case). The selection of diverse cases has the additional advantage of introducing variation on the key variables of interest. A set of diverse cases is, by definition, a set of cases that encompasses a range of high and low values on relevant dimensions. There is, therefore, much to recommend this method of case selection. I suspect that these advantages are commonly understood and are applied on an intuitive level by case-study researchers. However, the lack of a recognizable name—and an explicit methodological defense—has made it difficult for case-study researchers to utilize this method of case selection, and to do so in an explicit and self-conscious fashion. Neologism has its uses.

(p. 653) 3 Extreme Case

The *extreme-case* method selects a case because of its extreme value on an independent (X_1) or dependent (Y) variable of interest. Thus, studies of domestic violence may choose to focus on extreme instances of abuse (Browne 1987). Studies of altruism may focus on those rare individuals who risked their lives to help others (e.g. Holocaust resisters) (Monroe 1996). Studies of ethnic politics may focus on the most heterogeneous societies (e.g. Papua New Guinea) in order to better understand the role of ethnicity in a democratic setting (Reilly 2000–1). Studies of industrial policy often focus on the most successful countries (i.e. the NICS) (Deyo 1987). And so forth.[5]

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Often an extreme case corresponds to a case that is considered to be prototypical or paradigmatic of some phenomena of interest. This is because concepts are often defined by their extremes, i.e. their ideal types. Italian Fascism defines the concept of Fascism, in part, because it offered the most extreme example of that phenomenon. However, the *methodological* value of this case, and others like it, derives from its extremity (along some dimension of interest), not its theoretical status or its status in the literature on a subject.

The notion of "extreme" may now be defined more precisely. An extreme value is an observation that lies far away from the mean of a given distribution. This may be measured (if there are sufficient observations) by a case's "Z score"—the number of standard deviations between a case and the mean value for that sample. Extreme cases have high *Z* scores, and for this reason may serve as useful subjects for intensive analysis.

For a continuous variable, the distance from the mean may be in either direction (positive or negative). For a dichotomous variable (present/absent), extremeness may be interpreted as *unusual*. If most cases are positive along a given dimension, then a negative case constitutes an extreme case. If most cases are negative, then a positive case constitutes an extreme case. It should be clear that researchers are not simply concerned with cases where something "happened," but also with cases where something did not. It is the rareness of the value that makes a case valuable, in this context, not its positive or negative value.[6] Thus, if one is studying state capacity, a case of state failure is probably more informative than a case of state endurance simply because the former is more unusual. Similarly, if one is interested in incest taboos a culture where the incest taboo is absent or weak is probably more useful than a culture where it is present or strong. Fascism is more important than nonfascism. And so forth. There is a good reason, therefore, why case studies of revolution tend to focus on "revolutionary" cases. Theda Skocpol (1979) had much more to learn from France than from Austro-Hungary since France was more unusual than Austro-Hungary within the population of nation states that Skocpol was (p. 654) concerned to explain. The reason is quite simple: There are fewer revolutionary cases than nonrevolutionary cases; thus, the variation that we explore as a clue to causal relationships is encapsulated in these cases, against a background of nonrevolutionary cases.

Note that the extreme-case method of case selection appears to violate the social science folk wisdom warning us not to "select on the dependent variable."[7] Selecting cases on the dependent variable is indeed problematic if a number of cases are chosen, all of which lie on one end of a variable's spectrum (they are all positive *or* negative), and if the researcher then subjects this sample to cross-case analysis as if it were representative of a population.[8] Results for this sort of analysis would almost assuredly be biased. Moreover, there will be little variation to explain since the values of each case are explicitly constrained.

However, this is not the proper employment of the extreme-case method. (It is more appropriately labeled an extreme-*sample* method.) The extreme-case method actually

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

refers back to a larger sample of cases that lie in the background of the analysis and provide a full range of variation as well as a more representative picture of the population. It is a self-conscious attempt to *maximize* variance on the dimension of interest, not to minimize it. If this population of cases is well understood— either through the author's own cross-case analysis, through the work of others, or through common sense—then a researcher may justify the selection of a single case exemplifying an extreme value for within-case analysis. If not, the researcher may be well advised to follow a diverse-case method, as discussed above.

By way of conclusion, let us return to the problem of representativeness. It will be seen that an extreme case may be typical *or* deviant. There is simply no way to tell because the researcher has not yet specified an X_1/Y causal proposition. Once such a causal proposition has been specified one may then ask whether the case in question is similar to some population of cases in all respects that might affect the X_1/Y relationship of interest (i.e. unit homogeneous). It is at this point that it becomes possible to say, within the context of a cross-case statistical model, whether a case lies near to, or far from, the regression line. However, this sort of analysis means that the researcher is no longer pursuing an extreme-case method. The extreme-case method is purely exploratory—a way of probing possible causes of Y, or possible effects of X, in an open-ended fashion. If the researcher has some notion of what additional factors might affect the outcome of interest, or of what relationship the causal factor of interest might have with Y, then she ought to pursue one of the other methods explored in this chapter. This also implies that an extreme-case method may transform into a different kind of approach as a study evolves; that is, as a more specific hypothesis comes to light. Useful extreme cases at the outset of a study may prove less useful at a later stage of analysis.

(p. 655) 4 Deviant Case

The *deviant-case* method selects that case(s) which, by reference to some general understanding of a topic (either a specific theory or common sense), demonstrates a surprising value. It is thus the contrary of the typical case. Barbara Geddes (2003) notes the importance of deviant cases in medical science, where researchers are habitually focused on that which is "pathological" (according to standard theory and practice). The *New England Journal of Medicine*, one of the premier journals of the field, carries a regular feature entitled Case Records of the Massachusetts General Hospital. These articles bear titles like the following: "An 80-Year-Old Woman with Sudden Unilateral Blindness" or "A 76-Year-Old Man with Fever, Dyspnea, Pulmonary Infiltrates, Pleural Effusions, and Confusion."[9] Another interesting example drawn from the field of medicine concerns the extensive study now devoted to a small number of persons who seem resistant to the AIDS virus (Buchbinder and Vittinghoff 1999; Haynes, Pantaleo, and Fauci 1996). Why are they resistant? What is different about these people? What can we learn about AIDS in other patients by observing people who have built-in resistance to this disease?

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Likewise, in psychology and sociology case studies may be comprised of deviant (in the social sense) persons or groups. In economics, case studies may consist of countries or businesses that overperform (e.g. Botswana; Microsoft) or underperform (e.g. Britain through most of the twentieth century; Sears in recent decades) relative to some set of expectations. In political science, case studies may focus on countries where the welfare state is more developed (e.g. Sweden) or less developed (e.g. the United States) than one would expect, given a set of general expectations about welfare state development. The deviant case is closely linked to the investigation of theoretical anomalies. Indeed, to say deviant is to imply "anomalous."[10]

Note that while extreme cases are judged relative to the mean of a single distribution (the distribution of values along a single variable), deviant cases are judged relative to some general model of causal relations. The deviant-case method selects cases which, by reference to some (presumably) general relationship, demonstrate a surprising value. They are "deviant" in that they are poorly explained by the multivariate model. The important point is that deviant-ness can only be assessed relative to the general (quantitative or qualitative) model. This means that the relative deviant-ness of a case is likely to change whenever the general model is altered. For example, the United States is a deviant welfare state when this outcome is gauged relative to societal wealth. But it is less deviant—and perhaps not deviant at all—when certain additional (political and societal) factors are included in the model, as discussed in (p. 656) the epilogue. Deviance is model dependent. Thus, when discussing the concept of the deviant case it is helpful to ask the following question: *Relative to what general model* (or set of background factors) is Case A deviant?

Conceptually, we have said that the deviant case is the logical contrary of the typical case. This translates into a directly contrasting statistical measurement. While the typical case is one with a low residual (in some general model of causal relations), a deviant case is one with a high residual. This means, following our previous discussion, that the deviant case is likely to be an *un*representative case, and in this respect appears to violate the supposition that case-study samples should seek to reproduce features of a larger population.

However, it must be borne in mind that the primary purpose of a deviant-case analysis is to probe for new—but as yet unspecified—explanations. (If the purpose is to disprove an extant theory I shall refer to the study as crucial-case, as discussed below.) The researcher hopes that causal processes identified within the deviant case will illustrate some causal factor that is applicable to other (more or less deviant) cases. This means that a deviant-case study usually culminates in a general proposition, one that may be applied to other cases in the population. Once this general proposition has been introduced into the overall model, the expectation is that the chosen case will no longer be an outlier. Indeed, the hope is that it will now be *typical*, as judged by its small residual in the adjusted model. (The exception would be a circumstance in which a case's outcome is deemed to be "accidental," and therefore inexplicable by any general model.)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

This feature of the deviant-case study should help to resolve questions about its representativeness. Even if it is not possible to measure the new causal factor (and thus to introduce it into a large-*N* cross-case model), it may still be plausible to assert (based on general knowledge of the phenomenon) that the chosen case is representative of a broader population.

5 Influential Case

Sometimes, the choice of a case is motivated solely by the need to verify the assumptions behind a general model of causal relations. Here, the analyst attempts to provide a rationale for disregarding a problematic case or a set of problematic cases. That is to say, she attempts to show why apparent deviations from the norm are not *really* deviant, or do not challenge the core of the theory, once the circumstances of the special case or cases are fully understood. A cross-case analysis may, after all, be marred by several classes of problems including measurement error, specification error, errors in establishing proper boundaries for the inference (the scope of the argument), and stochastic error (fluctuations in the phenomenon under study that are treated as random, given available theoretical resources). If poorly fitting cases (p. 657) can be explained away by reference to these kinds of problems, then the theory of interest is that much stronger. This sort of deviant-case analysis answers the question, “What about Case A (or cases of type A)? How does that, seemingly disconfirming, case fit the model?”

Because its underlying purpose is different from the usual deviant-case study, I offer a new term for this method. The *influential case* is a case that casts doubt upon a theory, and for that reason warrants close inspection. This investigation may reveal, after all, that the theory is validated—perhaps in some slightly altered form. In this guise, the influential case is the “case that proves the rule.” In other instances, the influential-case analysis may contribute to disconfirming, or reconceptualizing, a theory. The key point is that the value of the case is judged relative to some extant cross-case model.

A simple version of influential-case analysis involves the confirmation of a key case's score on some critical dimension. This is essentially a question of measurement. Sometimes cases are poorly explained simply because they are poorly understood. A close examination of a particular context may reveal that an apparently falsifying case has been miscoded. If so, the initial challenge presented by that case to some general theory has been obviated.

However, the more usual employment of the influential-case method culminates in a substantive reinterpretation of the case—perhaps even of the general model. It is not just a question of measurement. Consider Thomas Ertman's (1997) study of state building in Western Europe, as summarized by Gerardo Munck. This study argues

> that the interaction of a) the type of local government during the first period of statebuilding, with b) the timing of increases in geopolitical competition, strongly influences the kind of regime and state that emerge. [Ertman] tests this

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

> hypothesis against the historical experience of Europe and finds that most countries fit his predictions. Denmark, however, is a major exception. In Denmark, sustained geopolitical competition began relatively late and local government at the beginning of the statebuilding period was generally participatory, which should have led the country to develop “patrimonial constitutionalism.” But in fact, it developed “bureaucratic absolutism.” Ertman carefully explores the process through which Denmark came to have a bureaucratic absolutist state and finds that Denmark had the early marks of a patrimonial constitutionalist state. However, the country was pushed off this developmental path by the influence of German knights, who entered Denmark and brought with them German institutions of local government. Ertman then traces the causal process through which these imported institutions pushed Denmark to develop bureaucratic absolutism, concluding that this development was caused by a factor well outside his explanatory framework.
>
> (Munck 2004, 118)

Ertman's overall framework is confirmed insofar as he has been able to show, by an in-depth discussion of Denmark, that the causal processes stipulated by the general theory hold even in this apparently disconfirming case. Denmark is still deviant, but it is so because of “contingent historical circumstances” that are exogenous to the theory (Ertman 1997, 316).
Evidently, the influential-case analysis is similar to the deviant-case analysis. Both focus on outliers. However, as we shall see, they focus on different kinds of outliers. (p. 658) Moreover, the animating goals of these two research designs are quite different. The influential-case study begins with the aim of confirming a general model, while the deviant-case study has the aim of generating a new hypothesis that modifies an existing general model. The confusion stems from the fact that the same case study may fulfill both objectives—qualifying a general model and, at the same time, confirming its core hypothesis.

Thus, in their study of Roberto Michels's “iron law of oligarchy,” Lipset, Trow, and Coleman (1956) choose to focus on an organization—the International Typographical Union—that appears to violate the central presupposition. The ITU, as noted by one of the authors, has “a long-term two-party system with free elections and frequent turnover in office” and is thus anything but oligarchic (Lipset 1959, 70). As such, it calls into question Michels's grand generalization about organizational behavior. The authors explain this curious result by the extraordinarily high level of education among the members of this union. Michels's law is shown to be true for most organizations, but not all. It is true, with qualifications. Note that the respecification of the original model (in effect, Lipset, Trow, and Coleman introduce a new control variable or boundary condition) involves the exploration of a new hypothesis. In this instance, therefore, the use of an influential case to confirm an existing theory is quite similar to the use of a deviant case to explore a new theory.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

In a quantitative idiom, influential cases are those that, if counterfactually assigned a different value on the dependent variable, would most substantially change the resulting estimates. They may or may not be outliers (high-residual cases). Two quantitative measures of influence are commonly applied in regression diagnostics (Belsey, Kuh, and Welsch 2004). The first, often referred to as the leverage of a case, derives from what is called the *hat matrix*. Based solely on each case's scores on the independent variables, the hat matrix tells us how much a change in (or a measurement error on) the dependent variable for that case would affect the overall regression line. The second is Cook's *distance*, a measure of the extent to which the estimates of all the parameters would change if a given case were omitted from the analysis. Cases with a large leverage or Cook's distance contribute quite a lot to the inferences drawn from a cross-case analysis. In this sense, such cases are vital for maintaining analytic conclusions. Discovering a significant measurement error on the dependent variable or an important omitted variable for such a case may dramatically revise estimates of the overall relationships. Hence, it may be quite sensible to select influential cases for in-depth study.

Note that the use of an influential-case strategy of case selection is limited to instances in which a researcher has reason to be concerned that her results are being driven by one or a few cases. This is most likely to be true in small to moderate-sized samples. Where N is very large—greater than 1,000, let us say—it is extremely unlikely that a small set of cases (much less an individual case) will play an “influential” role. Of course, there may be influential *sets* of cases, e.g. countries within a particular continent or cultural region, or persons of Irish extraction. Sets of influential observations are often problematic in a time-series cross-section data-set where each (p. 659) unit (e.g. country) contains multiple observations (through time), and hence may have a strong influence on aggregate results. Still, the general rule is: the larger the sample, the less important individual cases are likely to be and, hence, the less likely a researcher is to use an influential-case approach to case selection.

6 Crucial Case

Of all the extant methods of case selection perhaps the most storied—and certainly the most controversial—is the *crucial-case* method, introduced to the social science world several decades ago by Harry Eckstein. In his seminal essay, Eckstein (1975, 118) describes the crucial case as one “that *must closely fit* a theory if one is to have confidence in the theory's validity, or, conversely, *must not fit* equally well any rule contrary to that proposed.” A case is crucial in a somewhat weaker—but much more common—sense when it is most, or least, likely to fulfill a theoretical prediction. A “most-likely” case is one that, on all dimensions *except* the dimension of theoretical interest, is predicted to achieve a certain outcome, and yet does not. It is therefore used to disconfirm a theory. A “least-likely” case is one that, on all dimensions *except* the dimension of theoretical interest, is predicted not to achieve a certain outcome, and yet does so. It is therefore used to confirm a theory. In all formulations, the crucial-case

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

offers a most-difficult test for an argument, and hence provides what is perhaps the strongest sort of evidence possible in a nonexperimental, single-case setting.

Since the publication of Eckstein's influential essay, the crucial-case approach has been claimed in a multitude of studies across several social science disciplines and has come to be recognized as a staple of the case-study method.[11] Yet the idea of any single case playing a crucial (or "critical") role is not widely accepted among most methodologists (e.g. Sekhon 2004). (Even its progenitor seems to have had doubts.)

Let us begin with the confirmatory (a.k.a. least-likely) crucial case. The implicit logic of this research design may be summarized as follows. Given a set of facts, we are asked to contemplate the probability that a given theory is true. While the facts matter, to be sure, the effectiveness of this sort of research also rests upon the formal properties of the theory in question. Specifically, the degree to which a theory is amenable to confirmation is contingent upon how many predictions can be derived from the theory and on how "risky" each individual prediction is. In Popper's (1963, 36) words, "Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an (p. 660) event which was incompatible with the theory—and event which would have refuted the theory. Every 'good' scientific theory is a prohibition; it forbids certain things to happen. The more a theory forbids, the better it is" (see also Popper 1934/1968). A risky prediction is therefore one that is highly precise and determinate, and therefore unlikely to be achieved by the product of other causal factors (external to the theory of interest) or through stochastic processes. A theory produces many such predictions if it is fully elaborated, issuing predictions not only on the central outcome of interest but also on specific causal mechanisms, and if it is broad in purview. (The notion of riskiness may also be conceptualized within the Popperian lexicon as *degrees of falsifiability*.)

These points can also be articulated in Bayesian terms. Colin Howson and Peter Urbach explain: "The degree to which *h* [a hypothesis] is confirmed by *e* [a set of evidence] depends … on the extent to which *P(eČh)* exceeds *P (e)*, that is, on how much more probable *e* is relative to the hypothesis and background assumptions than it is relative just to background assumptions." Again, "confirmation is correlated with how much more probable the evidence is if the hypothesis is true than if it is false" (Howson and Urlbach 1989, 86). Thus, the stranger the prediction offered by a theory—relative to what we would normally expect—the greater the degree of confirmation that will be afforded by the evidence. As an intuitive example, Howson and Urbach (1989, 86) offer the following:

> If a soothsayer predicts that you will meet a dark stranger sometime and you do in fact, your faith in his powers of precognition would not be much enhanced: you would probably continue to think his predictions were just the result of guesswork. However, if the prediction also gave the correct number of hairs on the head of that stranger, your previous scepticism would no doubt be severely shaken.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

While these Popperian/Bayesian notions[12] are relevant to all empirical research designs, they are especially relevant to case-study research designs, for in these settings a single case (or, at most, a small number of cases) is required to bear a heavy burden of proof. It should be no surprise, therefore, that Popper's idea of “riskiness” was to be appropriated by case-study researchers like Harry Eckstein to validate the enterprise of single-case analysis. (Although Eckstein does not cite Popper the intellectual lineage is clear.) Riskiness, here, is analogous to what is usually referred to as a “most- difficult” research design, which in a case-study research design would be understood as a “least-likely” case. Note also that the distinction between a “must-fit” case and a least-likely case—that, in the event, actually does fit the terms of a theory—is a matter of degree. Cases are more or less crucial for confirming theories. The point is that, in some circumstances, a paucity of empirical evidence may be compensated by the riskiness of the theory.

The crucial-case research design is, perforce, a highly deductive enterprise; much depends on the quality of the theory under investigation. It follows that the theories most amenable to crucial-case analysis are those which are lawlike in their precision, (p. 661) degree of elaboration, consistency, and scope. The more a theory attains the status of a causal law, the easier it will be to confirm, or to disconfirm, with a single case. Indeed, risky predictions are common in natural science fields such as physics, which in turn served as the template for the deductive-nomological (“covering-law”) model of science that influenced Eckstein and others in the postwar decades (e.g. Hempel 1942).

A frequently cited example is the first important empirical demonstration of the theory of relativity, which took the form of a single-event prediction on the occasion of the May 29, 1919, solar eclipse (Eckstein 1975; Popper 1963). Stephen Van Evera (1997, 66–7) describes the impact of this prediction on the validation of Einstein's theory.

> Einstein's theory predicted that gravity would bend the path of light toward a gravity source by a specific amount. Hence it predicted that during a solar eclipse stars near the sun would appear displaced—stars actually behind the sun would appear next to it, and stars lying next to the sun would appear farther from it—and it predicted the amount of apparent displacement. No other theory made these predictions. The passage of this one single-case-study test brought the theory wide acceptance because the tested predictions were unique—there was no plausible competing explanation for the predicted result—hence the passed test was very strong.

The strength of this test is the extraordinary fit between the theory and a set of facts found in a single case, and the corresponding lack of fit between all other theories and this set of facts. Einstein offered an explanation of a particular set of anomalous findings that no other existing theory could make sense of. Of course, one must assume that there was no—or limited—measurement error. And one must assume that the phenomenon of interest is largely invariant; light does not bend differently at different times and places (except in ways that can be understood through the theory of relativity). And one must assume, finally, that the theory itself makes sense on other grounds (other than the case of special interest); it is a plausible general theory. If one is willing to accept these a priori assumptions, then the 1919 “case study” provides a very strong confirmation of the theory. It is difficult to imagine a stronger proof of the theory from within an observational (nonexperimental) setting.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

In social science settings, by contrast, one does not commonly find single-case studies offering knockout evidence for a theory. This is, in my view, largely a product of the looseness (the underspecification) of most social science theories. George and Bennett point out that while the thesis of the democratic peace is as close to a "law" as social science has yet seen, it cannot be confirmed (or refuted) by looking at specific causal mechanisms because the causal pathways mandated by the theory are multiple and diverse. Under the circumstances, no single-case test can offer strong confirmation of the theory (George and Bennett 2005, 209).

However, if one adopts a softer version of the crucial-case method—the least-likely (most difficult) case—then possibilities abound. Indeed, I suspect that, *implicitly*, most case-study work that makes a positive argument focusing on a single case (without a corresponding cross-case analysis) relies largely on the logic of the least- likely case. Rarely is this logic made explicit, except perhaps in a passing phrase or two. Yet the deductive logic of the "risky" prediction is central to the case-study enterprise. (p. 662) Whether a case study is convincing or not often rests on the reader's evaluation of how strong the evidence for an argument might be, and this in turn—wherever cross- case evidence is limited and no manipulated treatment can be devised—rests upon an estimation of the degree of "fit" between a theory and the evidence at hand, as discussed.

Lily Tsai's (2007) investigation of governance at the village level in China employs several in-depth case studies of villages which are chosen (in part) because of their least-likely status relative to the theory of interest. Tsai's hypothesis is that villages with greater social solidarity (based on preexisting religious or familial networks) will develop a higher level of social trust and mutual obligation and, as a result, will experience better governance. Crucial cases, therefore, are villages that evidence a high level of social solidarity but which, along other dimensions, would be judged least likely to develop good governance, e.g. they are poor, isolated, and lack democratic institutions or accountability mechanisms from above. "Li Settlement," in Fujian province, is such a case. The fact that this impoverished village nonetheless boasts an impressive set of infrastructural accomplishments such as paved roads with drainage ditches (a rarity in rural China) suggests that something rather unusual is going on here. Because her case is carefully chosen to eliminate rival explanations, Tsai's conclusions about the special role of social solidarity are difficult to gainsay. How else is one to explain this otherwise anomalous result? This is the strength of the least-likely case, where all other plausible causal factors for an outcome have been minimized.[13]

Jack Levy (2002, 144) refers to this, evocatively, as a "Sinatra inference:" if it can make it here, it can make it anywhere (see also Khong 1992, 49; Sagan 1995, 49; Shafer 1988, 14–6). Thus, if social solidarity has the hypothesized effect in Li Settlement it should have the same effect in more propitious settings (e.g. where there is greater economic surplus). The same implicit logic informs many case-study analyses where the intent of the study is to confirm a hypothesis on the basis of a single case.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Another sort of crucial case is employed for the purpose of *dis*confirming a causal hypothesis. A central Popperian insight is that it is easier to disconfirm an inference than to confirm that same inference. (Indeed, Popper doubted that any inference could be fully confirmed, and for this reason preferred the term "corroborate.") This is particularly true of case-study research designs, where evidence is limited to one or several cases. The key proviso is that the theory under investigation must take a consistent (a.k.a. invariant, deterministic) form, even if its predictions are not terrifically precise, well elaborated, or broad.

As it happens, there are a fair number of invariant propositions floating around the social science disciplines (Goertz and Levy forthcoming; Goertz and Starr 2003). It used to be argued, for example, that political stability would occur only in countries that are relatively homogeneous, or where existing heterogeneities are mitigated by (p. 663) cross-cutting cleavages (Almond 1956; Bentley 1908/1967; Lipset 1960/1963; Truman 1951). Arend Lijphart's (1968) study of the Netherlands, a peaceful country with reinforcing social cleavages, is commonly viewed as refuting this theory on the basis of a single in-depth case analysis.[14]

Granted, it may be questioned whether presumed invariant theories are *really* invariant; perhaps they are better understood as probabilistic. Perhaps, that is, the theory of cross-cutting cleavages is still true, probabilistically, despite the apparent Dutch exception. Or perhaps the theory is still true, deterministically, within a subset of cases that does not include the Netherlands. (This sort of claim seems unlikely in this particular instance, but it is quite plausible in many others.) Or perhaps the theory is in need of reframing; it is true, deterministically, but applies only to cross- cutting *ethnic/racial* cleavages, not to cleavages that are primarily religious. One can quibble over what it means to "disconfirm" a theory. The point is that the crucial case has, in all these circumstances, provided important updating of a theoretical prior.

Heretofore, I have treated causal factors as dichotomous. Countries have either reinforcing or cross-cutting cleavages and they have regimes that are either peaceful or conflictual. Evidently, these sorts of parameters are often matters of degree. In this reading of the theory, cases are *more or less* crucial. Accordingly, the most useful—i.e. most crucial—case for Lijphart's purpose is one that has the most segregated social groups and the most peaceful and democratic track record. In these respects, the Netherlands was a very good choice. Indeed, the degree of disconfirmation offered by this case study is probably greater than the degree of disconfirmation that might have been provided by other cases such as India or Papua New Guinea—countries where social peace has not always been secure. The point is that where variables are continuous rather than dichotomous it is possible to evaluate potential cases in terms of their *degree of crucialness*.

Note that the crucial-case method of case-selection, whether employed in a confirmatory or disconfirmatory mode, cannot be employed in a large-*N* context. This is because an explicit cross-case model would render the crucial-case study redundant. Once one

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

identifies the relevant parameters and the scores of all cases on those parameters, one has in effect constructed a cross-case model that confirms or disconfirms the theory in question. The case study is thenceforth irrelevant, at least as a means of decisive confirmation or disconfirmation.[15] It remains highly relevant as a means of exploring causal mechanisms, of course. Yet, because this objective is quite different from that which is usually associated with the term, I enlist a new term for this technique.

(p. 664) 7 Pathway Case

One of the most important functions of case-study research is the elucidation of causal mechanisms. But which sort of case is most useful for this purpose? Although all case studies presumably shed light on causal mechanisms, not all cases are equally transparent. In situations where a causal hypothesis is clear and has already been confirmed by cross-case analysis, researchers are well advised to focus on a case where the causal effect of X_1 on Y can be isolated from other potentially confounding factors (X_2). I shall call this a *pathway case* to indicate its uniquely penetrating insight into causal mechanisms. In contrast to the crucial case, this sort of method is practicable only in circumstances where cross-case covariational patterns are well studied and where the mechanism linking X_1 and Y remains dim. Because the pathway case builds on prior cross-case analysis, the problem of case selection must be situated within that sample. There is no standalone pathway case.

The logic of the pathway case is clearest in situations of causal sufficiency—where a causal factor of interest, X_1, is sufficient by itself (though perhaps not necessary) to account for Y's value (0 or 1). The other causes of Y, about which we need make no assumptions, are designated as a vector, X_2.

Note that wherever various causal factors are substitutable for one another, each factor is conceptualized (individually) as sufficient (Braumoeller 2003). Thus, situations of causal equifinality presume causal sufficiency on the part of each factor or set of conjoint factors. An example is provided by the literature on democratization, which stipulates three main avenues of regime change: leadership-initiated reform, a controlled opening to opposition, or the collapse of an authoritarian regime (Colomer 1991). The case-study format constrains us to analyze one at a time, so let us limit our scope to the first one—leadership-initiated reform. So considered, a causal-pathway case would be one with the following features: (a) democratization, (b) leadership-initiated reform, (c) no controlled opening to the opposition, (d) no collapse of the previous authoritarian regime, and (e) no other extraneous factors that might affect the process of democratization. In a case of this type, the causal mechanisms by which leadership-initiated reform may lead to democratization will be easiest to study. Note that it is not necessary to assume that leadership-initiated reform *always* leads to democratization; it may or may not be a deterministic cause. But it is necessary to assume that leadership-initiated reform can *sometimes* lead to democratization on its own (given certain background features).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Now let us move from these examples to a general-purpose model. For heuristic purposes, let us presume that all variables in that model are dichotomous (coded as 0 or 1) and that the model is complete (all causes of Y are included). All causal relationships will be coded so as to be positive: X_1 and Y covary as do X_2 and Y. This allows us to visualize a range of possible combinations at a glance.

Recall that the pathway case is always focused, by definition, on a single causal factor, denoted X_1. (The researcher's focus may shift to other causal factors, but may only focus on one causal factor at a time.) In this scenario, and regardless of how (p. 665) many additional causes of Y there might be (denoted X_2, a vector of controls), there are only eight *relevant* case types, as illustrated in Table 28.2. Identifying these case types is a relatively simple matter, and can be accomplished in a small-N sample by the construction of a truth-table (modeled after Table 28.2) or in a large-N sample by the use of cross-tabs.

Table 28.2. Pathway case with dichotomous causal factors

Case types	X_1	X_2	Y
A	1	1	1
B	0	0	0
C	0	1	1
D	0	0	1
E	1	0	0
F	1	1	0
G	1	0	1
H	0	1	0

Notes: X_1 = the variable of theoretical interest. X_2 = a vector of controls (a score of 0 indicates that all control variables have a score of 0, while a score of 1 indicates that all control variables have a score of 1). Y = the outcome of interest. A–H = case types (the N for each case type is indeterminate). G, H = possible pathway cases. Sample size = indeterminate.

Assumptions: (a) all variables can be coded dichotomously (a binary coding of the concept is valid); (b) all independent variables are positively correlated with Y in the general case; (*c*)X_1 is (at least sometimes) a sufficient cause of Y.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Note that the total number of combinations of values depends on the number of control variables, which we have represented with a single vector, X_2. If this vector consists of a single variable then there are only eight case types. If this vector consists of two variables (X_{2a}, X_{2b}) then the total number of possible combinations increases from eight (2^3) to sixteen (2^4). And so forth. However, none of these combinations is relevant for present purposes except those where X_{2a} and X_{2b} have the *same value* (0 or 1). "Mixed" cases are not causal pathway cases, for reasons that should become clear.

The pathway case, following the logic of the crucial case, is one where the causal factor of interest, X_1, correctly predicts Y while all other possible causes of Y (represented by the vector, X_2) make "wrong" predictions. If X_1 is—at least in some circumstances—a sufficient cause of Y, then it is these sorts of cases that should be most useful for tracing causal mechanisms. There are only two such cases in Ta b l e 28.2—G and H. In all other cases, the mechanism running from X_1 to Y would be difficult to discern either because X_1 and Y are not correlated in the usual way (constituting an unusual case, in the terms of our hypothesis) or because other confounding factors (X_2) intrude. In case A, for example, the positive value on Y (p. 666) could be a product of X_1 or X_2. An in-depth examination of this case is not likely to be very revealing.

Keep in mind that because the researcher already knows from her cross-case examination what the general causal relationships are, she knows (prior to the case- study investigation) what constitutes a correct or incorrect prediction. In the crucial- case method, by contrast, these expectations are deductive rather than empirical. This is what differentiates the two methods. And this is why the causal pathway case is useful principally for elucidating causal mechanisms rather than verifying or falsifying general propositions (which are already more or less apparent from the cross-case evidence). Of course, we must leave open the possibility that the investigation of causal mechanisms would invalidate a general claim, if that claim is utterly contingent upon a specific set of causal mechanisms and the case study shows that no such mechanisms are present. However, this is rather unlikely in most social science settings. Usually, the result of such a finding will be a reformulation of the causal processes by which X_1 causes Y—or, alternatively, a realization that the case under investigation is aberrant (atypical of the general population of cases).

Sometimes, the research question is framed as a unidirectional cause: one is interested in why 0 becomes 1 (or vice versa) but not in why 1 becomes 0. In our previous example, we asked why democracies fail, *not* why countries become democratic or authoritarian. So framed, there can be only one type of causal-pathway case. (Whether regime failure is coded as 0 or 1 is a matter of taste.) Where researchers are interested in bidirectional causality—a movement from 0 to 1 as well as from 1 to 0—there are two possible causal-pathway cases, G and H. In practice, however, one of these case types is almost always more useful than the other. Thus, it seems reasonable to employ the term "pathway case" in the singular. In order to determine which of these two case types will be more useful for intensive analysis the researcher should look to see whether each case type exhibits desirable features such as: (a) a rare (unusual) value on X_1 or Y (designated "extreme" in

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

our previous discussion), (b) observable temporal variation in X_1, (*c*) an X_1/Y relationship that is easier to study (it has more visible features; it is more transparent), or (d) a lower residual (thus indicating a more typical case, within the terms of the general model). Usually, the choice between G and H is intuitively obvious.

Now, let us consider a scenario in which all (or most) variables of concern to the model are continuous, rather than dichotomous. Here, the job of case selection is considerably more complex, for causal "sufficiency" (in the usual sense) cannot be invoked. It is no longer plausible to assume that a given cause can be entirely partitioned, i.e. rival factors eliminated. However, the search for a pathway case may still be viable. What we are looking for in this scenario is a case that satisfies two criteria: (1) it is not an outlier (or at least not an extreme outlier) in the general model and (2) its score on the outcome (Y) is strongly influenced by the theoretical variable of interest (X_1), taking all other factors into account (X_2). In this sort of case it should be easiest to "see" the causal mechanisms that lie between X_1 and Y.

Achieving the second desiderata requires a bit of manipulation. In order to determine which (nonoutlier) cases are most strongly affected by X_1, given all the other (p. 667) parameters in the model, one must compare the size of the residuals for each case in a reduced form model, $Y = \text{Constant} + X_2 + \text{Res}_{\text{reduced}}$, with the size of the residuals for each case in a full model, $Y = \text{Constant} + X_2 + X_1 + \text{Res}_{\text{full}}$. The pathway case is that case, or set of cases, which shows the greatest difference between the residual for the reduced-form model and the full model (ΔResidual). Thus,

$$\text{Pathway} = |\text{Res}_{\text{reduced}} - \text{Res}_{\text{full}}|, \text{ if } |\text{Res}_{\text{reduced}}| > |\text{Res}_{\text{full}}|. \quad (1)$$

Note that the residual for a case must be smaller in the full model than in the reduced- form model; otherwise, the addition of the variable of interest (X_1) pulls the case *away* from the regression line. We want to find a case where the addition of X_1 pushes the case towards the regression line, i.e. it helps to "explain" that case.

As an example, let us suppose that we are interested in exploring the effect of mineral wealth on the prospects for democracy in a society. According to a good deal of work on this subject, countries with a bounty of natural resources—particularly oil—are less likely to democratize (or once having undergone a democratic transition, are more likely to revert to authoritarian rule) (Barro 1999; Humphreys 2005; Ross 2001). The cross-country evidence is robust. Yet as is often the case, the causal mechanisms remain rather obscure. In order to better understand this phenomenon it may be worthwhile to exploit the findings of cross-country regression models in order to identify a country whose regime type (i.e. its democracy "score" on some general index) is strongly affected by its natural-research wealth, all other things held constant. An analysis of this sort identifies two countries— the United Arab Emirates and Kuwait—with high Δ Residual values and modest residuals in the full model (signifying that these cases are not outliers). Researchers seeking to explore the effect of oil wealth on regime type might do well to focus on these two cases since their patterns of democracy cannot be well explained by other factors—e.g. economic development, religion, European influence, or ethnic fractionalization. The presence of oil wealth in these countries would appear to have a

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

strong *independent* effect on the prospects for democratization in these cases, an effect that is well modeled by general theory and by the available cross-case evidence.

To reiterate, the logic of causal "elimination" is much more compelling where variables are dichotomous and where causal sufficiency can be assumed (X_1 is sufficient by itself, at least in some circumstances, to cause Y). Where variables are continuous, the strategy of the pathway case is more dubious, for potentially confounding causal factors (X_2) cannot be neatly partitioned. Even so, we have indicated why the selection of a pathway case may be a logical approach to case-study analysis in many circumstances.

The exceptions may be briefly noted. Sometimes, where all variables in a model are dichotomous, there are no pathway cases, i.e. no cases of type G or H (in Table 28.2). This is known as the "empty cell" problem, or a problem of severe causal multicollinearity. The universe of observational data does not always oblige us with cases that allow us to independently test a given hypothesis. Where variables are continuous, the analogous problem is that of a causal variable of interest (X_1) that (p. 668) has only minimal effects on the outcome of interest. That is, its role in the general model is quite minor. In these situations, the only cases that are strongly affected by X_1—if there are any at all—may be extreme outliers, and these sorts of cases are not properly regarded as providing confirmatory evidence for a proposition, for reasons that are abundantly clear by now.

Finally, it should be clarified that the identification of a causal pathway case does not obviate the utility of exploring other cases. One might, for example, want to compare both sorts of potential pathway cases—G *and* H—with each other. Many other combinations suggest themselves. However, this sort of multi-case investigation moves beyond the logic of the causal-pathway case.

8 Most-similar Cases

The most-similar method employs a minimum of two cases.[16] In its purest form, the chosen pair of cases is similar in all respects except the variable(s) of interest. If the study is *exploratory* (i.e. hypothesis generating), the researcher looks for cases that differ on the outcome of theoretical interest but are similar on various factors that might have contributed to that outcome, as illustrated in Table 28.3 (A). This is a common form of case selection at the initial stage of research. Often, fruitful analysis begins with an apparent anomaly: two cases are apparently quite similar, and yet demonstrate surprisingly different outcomes. The hope is that intensive study of these cases will reveal one—or at most several—factors that differ across these cases. These differing factors (X_1) are looked upon as putative causes. At this stage, the research may be described by the second diagram in Table 28.3 (B). Sometimes, a researcher begins with a strong hypothesis, in which case her research design is confirmatory (hypothesis testing) from the get-go. That is, she strives to identify cases that exhibit different outcomes, different scores on the factor of interest, and similar scores on all other possible causal factors, as illustrated in the second (hypothesis-testing) diagram in Table 28.3 (B).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

The point is that the purpose of a most-similar research design, and hence its basic setup, often changes as a researcher moves from an exploratory to a confirmatory mode of analysis. However, regardless of where one begins, the results, when published, look like a hypothesis-testing research design. Question marks have been removed: (A) becomes (B) in Table 28.3.

As an example, let us consider Leon Epstein's classic study of party cohesion, which focuses on two “most-similar” countries, the United States and Canada. Canada has highly disciplined parties whose members vote together on the floor of the House of (p. 669) Commons while the United States has weak, undisciplined parties, whose members often defect on floor votes in Congress. In explaining these divergent outcomes, persistent over many years, Epstein first discusses possible causal factors that are held more or less constant across the two cases. Both the United States and Canada inherited English political cultures, both have large territories and heterogeneous populations, both are federal, and both have fairly loose party structures with strong regional bases and a weak center. These are the “control” variables. Where they differ is in one constitutional feature: Canada is parliamentary while the United States is presidential. And it is this institutional difference that Epstein identifies as the crucial (differentiating) cause. (For further examples of the most-similar method see Brenner 1976; Hamilton 1977; Lipset 1968; Miguel 2004; Moulder 1977; Posner 2004.)

Table 28.3. Most-similar analysis with two case types

Case types	X_1	X_2	Y
(A) Hypothesis-generating (Y-centered):			
A	?	0	1
B	?	0	0
(B) Hypothesis-testing(X_1/*Y-centered*):			
A	1	0	1
B	0	0	0

X_1 = the variable of theoretical interest. X_2 = a vector of controls. Y = the outcome of interest.

Several caveats apply to any most-similar analysis (in addition to the usual set of assumptions applying to all case-study analysis). First, each causal factor is understood as having an independent and additive effect on the outcome; there are no “interaction” effects. Second, one must code cases dichotomously (high/low, present/absent). This is straightforward if the underlying variables are also dichotomous (e.g. federal/unitary).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

However, it is often the case that variables of concern in the model are continuous (e.g. party cohesion). In this setting, the researcher must “dichotomize” the scoring of cases so as to simplify the two-case analysis. (Some flexibility is admissible on the vector of controls (X_2) that are “held constant” across the cases. Nonidentity is tolerable if the deviation runs counter to the predicted hypothesis. For example, Epstein describes both the United States and Canada as having strong regional bases of power, a factor that is probably more significant in recent Canadian history than in recent American history. However, because regional bases of power should lead to weaker parties, rather than stronger parties, this element of nonidentity does not challenge Epstein's conclusions. Indeed, it sets up a most-difficult research scenario, as discussed above.)

In one respect the requirements for case control are not so stringent. Specifically, it is not usually necessary to *measure* control variables (at least not with a high degree of precision) in order to control for them. If two countries can be assumed (p. 670) to have similar cultural heritages one needn't worry about constructing variables to measure that heritage. One can simply assert that, whatever they are, they are more or less constant across the two cases. This is similar to the technique employed in a randomized experiment, where the researcher typically does not attempt to measure all the factors that might affect the causal relationship of interest. She assumes, rather, that these unknown factors have been neutralized across the treatment and control groups by randomization or by the choice of a sample that is internally homogeneous.

The most useful statistical tool for identifying cases for in-depth analysis in a most-similar setting is probably some variety of matching strategy—e.g. exact matching, approximate matching, or propensity-score matching.[17] The product of this procedure is a set of matched cases that can be compared in whatever way the researcher deems appropriate. These are the “most-similar” cases. Rosenbaum and Silber (2001, 223) summarize:

> Unlike model-based adjustments, where [individuals] vanish and are replaced by the coefficients of a model, in matching, ostensibly comparable patterns are compared directly, one by one. Modern matching methods involve statistical modeling and combinatorial algorithms, but the end result is a collection of pairs or sets of people who look comparable, at least on average. In matching, people retain their integrity as people, so they can be examined and their stories can be told individually.

Matching, conclude the authors, “facilitates, rather than inhibits, thick description” (Rosenbaum and Silber 2001, 223).

In principle, the same matching techniques that have been used successfully in observational studies of medical treatments might also be adapted to the study of nation states, political parties, cities, or indeed any traditional paired cases in the social sciences. Indeed, the current popularity of matching among statisticians—relative, that is, to garden-variety regression models—rests upon what qualitative researchers would recognize as a “case-based” approach to causal analysis. If Rosenbaum and Silber are

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

correct, it may be perfectly reasonable to appropriate this large-*N* method of analysis for case-study purposes.

As with other methods of case selection, the most-similar method is prone to problems of nonrepresentativeness. If employed in a qualitative fashion (without a systematic cross-case selection strategy), potential biases in the chosen case must be addressed in a speculative way. If the researcher employs a matching technique of case selection within a large-*N* sample, the problem of potential bias can be addressed by assuring the choice of cases that are not extreme outliers, as judged by their residuals in the full model. Most-similar cases should also be "typical" cases, though some scope for deviance around the regression line may be acceptable for purposes of finding a good fit among cases. (p. 671)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Table 28.4. Most-different analysis with two cases

Case types	X_1	X_{2a}	X_{2b}	X_{2c}	X_{2d}	Y
A	1	1	0	1	0	1
B	1	0	1	0	1	1

X_1 = the variable of theoretical interest. $X_{2a–d}$ = a vector of controls. Y = the outcome of interest.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

9 Most-different Cases

A final case-selection method is the reverse image of the previous method. Here, variation on independent variables is prized, while variation on the outcome is eschewed. Rather than looking for cases that are most-similar, one looks for cases that are *most- different*. Specifically, the researcher tries to identify cases where just one independent variable (X_1), as well as the dependent variable (Y), covary, while all other plausible factors ($X_{2a–d}$) show different values.[18]

The simplest form of this two-case comparison is illustrated in Table 28.4. Cases A and B are deemed “most different,” though they are similar in two essential respects— the causal variable of interest and the outcome.

As an example, I follow Marc Howard's (2003) recent work, which explores the enduring impact of Communism on civil society.[19] Cross-national surveys show a strong correlation between former Communist regimes and low social capital, controlling for a variety of possible confounders. It is a strong result. Howard wonders why this relationship is so strong and why it persists, and perhaps even strengthens, in countries that are no longer socialist or authoritarian. In order to answer this question, he focuses on two most-different cases, Russia and East Germany. These two countries were quite different—in all ways *other than* their Communist experience— prior to the Soviet era, during the Soviet era (since East Germany received substantial subsidies from West Germany), and in the post-Soviet era, as East Germany was absorbed into West Germany. Yet, they both score near the bottom of various cross- national indices intended to measure the prevalence of civic engagement in the current era. Thus, Howard's (2003, 6–9) case selection procedure meets the requirements of the most-different research design: Variance is found on all (or most) dimensions (p. 672) aside from the key factor of interest (Communism) and the outcome (civic engagement).

What leverage is brought to the analysis from this approach? Howard's case studies combine evidence drawn from mass surveys and from in-depth interviews of small, stratified samples of Russians and East Germans. (This is a good illustration, incidentally, of how quantitative and qualitative evidence can be fruitfully combined in the intensive study of several cases.) The product of this analysis is the identification of three causal pathways that, Howard (2003, 122) claims, help to explain the laggard status of civil society in post-Communist polities: “the mistrust of communist organizations, the persistence of friendship networks, and the disappointment with post-communism.” Simply put, Howard (2003, 145) concludes, “a great number of citizens in Russia and Eastern Germany feel a strong and lingering sense of distrust of any kind of public organization, a general satisfaction with their own personal networks (accompanied by a sense of deteriorating relations within society overall), and disappointment in the developments of post-communism.”

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

The strength of this most-different case analysis is that the results obtained in East Germany and Russia should also apply in other post-Communist polities (e.g. Lithuania, Poland, Bulgaria, Albania). By choosing a heterogeneous sample, Howard solves the problem of representativeness in his restricted sample. However, this sample is demonstrably *not* representative across the population of the inference, which is intended to cover all countries of the world.

More problematic is the lack of variation on key causal factors of interest— Communism and its putative causal pathways. For this reason, it is difficult to reach conclusions about the causal status of these factors on the basis of the most-different analysis alone. It is possible, that is, that the three causal pathways identified by Howard also operate within polities that never experienced Communist rule.

Nor does it seem possible to conclusively eliminate rival hypotheses on the basis of this most-different analysis. Indeed, this is not Howard's intention. He wishes merely to show that whatever influence on civil society might be attributed to economic, cultural, and other factors does not exhaust this subject.

My considered judgment is that the most-different research design provides minimal leverage into the problem of why Communist systems appear to suppress civic engagement, years after their disappearance. Fortunately, this is not the only research design employed by Howard in his admirable study. Indeed, the author employs two other small-*N* cross-case methods, as well as a large-*N* cross-country statistical analysis. These methods do most of the analytic work. East Germany may be regarded as a causal pathway case (see above). It has all the attributes normally assumed to foster civic engagement (e.g. a growing economy, multiparty competition, civil liberties, a free press, close association with Western European culture and politics), but nonetheless shows little or no improvement on this dimension during the post- transition era (Howard 2003, 8). It is plausible to attribute this lack of change to its Communist past, as Howard does, in which case East Germany should be a fruitful case for the investigation of causal mechanisms. The contrast between East and West Germany provides a most-similar analysis since the two polities share (p. 673) virtually everything except a Communist past. This variation is also deftly exploited by Howard.

I do not wish to dismiss the most-different research method entirely. Surely, Howard's findings are stronger with the intensive analysis of Russia than they would be without. Yet his book would not stand securely on the empirical foundation provided by most-different analysis alone. If one strips away the pathway-case (East Germany) and the most-similar analysis (East/West Germany) there is little left upon which to base an analysis of causal relations (aside from the large-*N* cross-national analysis). Indeed, most scholars who employ the most-different method do so in conjunction with other methods.[20] It is rarely, if ever, a standalone method.[21]

Generalizing from this discussion of Marc Howard's work, I offer the following summary remarks on the most-different method of case analysis. (I leave aside issues faced by *all* case-study analyses, issues that are explored in Gerring 2007.)

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Let us begin with a methodological obstacle that is faced by both Millean styles of analysis—the necessity of dichotomizing every variable in the analysis. Recall that, as with most-similar analysis, differences across cases must generally be sizeable enough to be interpretable in an essentially dichotomous fashion (e.g. high/low, present/absent) and similarities must be close enough to be understood as essentially identical (e.g. high/high, present/present). Otherwise the results of a Millean style analysis are not interpretable. The problem of "degrees" is deadly if the variables under consideration are, by nature, continuous (e.g. GDP). This is a particular concern in Howard's analysis, where East Germany scores somewhat higher than Russia in civic engagement; they are both low, but Russia is quite a bit lower. Howard assumes that this divergence is minimal enough to be understood as a difference of degrees rather than of kinds, a judgment that might be questioned. In these respects, most-different analysis is no more secure—but also no less—than most-similar analysis.

In one respect, most-different analysis is superior to most-similar analysis. If the coding assumptions are sound, the most-different research design may be quite useful for *eliminating necessary causes*. Causal factors that do not appear across the chosen cases—e.g. $X_{2a–d}$ in Table 28.4—are evidently unnecessary for the production of Y. However, it does not follow that the most-different method is the *best* method for eliminating necessary causes. Note that the defining feature of this method is the (p. 674) shared element across cases—X_1 in Table 28.4. This feature does not help one to eliminate necessary causes. Indeed, if one were focused solely on eliminating necessary causes one would presumably seek out cases that register the same outcomes and have maximum diversity on other attributes. In Table 28.4, this would be a set of cases that satisfy conditions $X_{2a–d}$, but not X_1. Thus, even the presumed strength of the most-different analysis is not so strong.

Usually, case-study analysis is focused on the identification (or clarification) of causal relations, not the elimination of possible causes. In this setting, the most- different technique is useful, but only if assumptions of causal uniqueness hold. By "causal uniqueness," I mean a situation in which a given outcome is the product of only one cause: Y cannot occur except in the presence of X. X is necessary, and in some situations (given certain background conditions) sufficient, to cause Y.[22]

Consider the following hypothetical example. Suppose that a new disease, about which little is known, has appeared in Country A. There are hundreds of infected persons across dozens of affected communities in that country. In Country B, located at the other end of the world, several new cases of the disease surface in a single community. In this setting, we can imagine two sorts of Millean analyses. The first examines two similar communities within Country A, one of which has developed the disease and the other of which has not. This is the most-similar style of case comparison, and focuses accordingly on the identification of a difference between the two cases that might account for variation across the sample. A second approach focuses on communities where the disease has

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

appeared across the two countries and searches for any similarities that might account for these similar outcomes. This is the most-different research design.

Both are plausible approaches to this particular problem, and we can imagine epidemiologists employing them simultaneously. However, the most-different design demands stronger assumptions about the underlying factors at work. It supposes that the disease arises from the *same cause* in any setting. This is often a reasonable operating assumption when one is dealing with natural phenomena, though there are certainly many exceptions. Death, for example, has many causes. For this reason, it would not occur to us to look for most-different cases of high mortality around the world. In order for the most-different research design to effectively identify a causal factor at work in a given outcome, the researcher must assume that X_1—the factor held constant across the diverse cases—is the only possible cause of Y (see Table 28.4). This assumption rarely holds in social-scientific settings. Most outcomes of interest to anthropologists, economists, political scientists, and sociologists have *multiple* causes. There are many ways to win an election, to build a welfare state, to get into a war, to overthrow a government, or—returning to Marc Howard's work—to build a strong civil society. And it is for this reason that most-different analysis is rarely applied in social science work and, where applied, is rarely convincing.

If this seems a tad severe, there is a more charitable way of approaching the most-different method. Arguably, this is not a pure “method” at all but merely a (p. 675) supplement, a way of incorporating diversity in the sub-sample of cases that provide the unusual outcome of interest. If the unusual outcome is revolutions, one might wish to encompass a wide variety of revolutions in one's analysis. If the unusual outcome is post-Communist civil society, it seems appropriate to include a diverse set of post-Communist polities in one's sample of case studies, as Marc Howard does. From this perspective, the most-different method (so-called) might be better labeled a *diverse-case* method, as explored above.

10 Conclusions

In order to be a case of something broader than itself, the chosen case must be representative (in some respects) of a larger population. Otherwise—if it is purely idiosyncratic (“unique”)—it is uninformative about anything lying outside the borders of the case itself. A study based on a nonrepresentative sample has no (or very little) external validity. To be sure, no phenomenon is purely idiosyncratic; the notion of a unique case is a matter that would be difficult to define. One is concerned, as always, with matters of degree. Cases are *more or less* representative of some broader phenomenon and, on that score, may be considered better or worse subjects for intensive analysis. (The one exception, as noted, is the influential case.)

Of all the problems besetting case-study analysis, perhaps the most persistent— and the most persistently bemoaned—is the problem of sample bias (Achen and Snidal 1989; Collier and Mahoney 1996; Geddes 1990; King, Keohane, and Verba 1994; Rohlfing 2004;

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

Sekhon 2004). Lisa Martin (1992, 5) finds that the overemphasis of international relations scholars on a few well-known cases of economic sanctions— most of which failed to elicit any change in the sanctioned country—"has distorted analysts view of the dynamics and characteristics of economic sanctions." Barbara Geddes (1990) charges that many analyses of industrial policy have focused exclusively on the most successful cases—primarily the East Asian NICs—leading to biased inferences. Anna Breman and Carolyn Shelton (2001) show that case-study work on the question of structural adjustment is systematically biased insofar as researchers tend to focus on disaster cases—those where structural adjustment is associated with very poor health and human development outcomes. These cases, often located in sub-Saharan Africa, are by no means representative of the entire population. Consequently, scholarship on the question of structural adjustment is highly skewed in a particular ideological direction (against neoliberalism) (see also Gerring, Thacker, and Moreno 2005).

These examples might be multiplied many times. Indeed, for many topics the most-studied cases are acknowledged to be less than representative. It is worth reflecting upon the fact that our knowledge of the world is heavily colored by a few "big" (populous, rich, powerful) countries, and that a good portion of the disciplines of economics, political science, and sociology are built upon scholars' familiarity with (p. 676) the economics, political science, and sociology of one country, the United States.[23] Case-study work is particularly prone to problems of investigator bias since so much rides on the researcher's selection of one (or a few) cases. Even if the investigator is unbiased, her sample may still be biased simply by virtue of "random" error (which may be understood as measurement error, error in the data-generation process, or as an underlying causal feature of the universe).

There are only two situations in which a case-study researcher need not be concerned with the representativeness of her chosen case. The first is the influential case research design, where a case is chosen because of its possible influence on a cross-case model, and hence is not expected to be representative of a larger sample. The second is the deviant-case method, where the chosen case is employed to confirm a broader cross-case argument to which the case stands as an apparent exception. Yet even here the chosen case is expected to be representative of a broader set of cases—those, in particular, that are poorly explained by the extant model.

In all other circumstances, cases must be representative of the population of interest in whatever ways might be relevant to the proposition in question. Note that where a researcher is attempting to disconfirm a deterministic proposition the question of representativeness is perhaps more appropriately understood as a question of classification: Is the chosen case appropriately classified as a member of the designated population? If so, then it is fodder for a disconfirming case study.

If the researcher is attempting to confirm a deterministic proposition, or to make probabilistic arguments about a causal relationship, then the problem of representativeness is of the more usual sort: Is case A unit-homogeneous relative to other

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

cases in the population? This is not an easy matter to test. However, in a large-*N* context the residual for that case (in whatever model the researcher has greatest confidence in) is a reasonable place to start. Of course, this test is only as good as the model at hand. Any incorrect specifications or incorrect modeling procedures will likely bias the results and give an incorrect assessment of each case's "typicality." In addition, there is the possibility of stochastic error, errors that cannot be modeled in a general framework. Given the explanatory weight that individual cases are asked to bear in a case-study analysis, it is wise to consider more than just the residual test of representativeness. Deductive logic and an in-depth knowledge of the case in question are often more reliable tools than the results of a cross-case model.

In any case, there is no dispensing with the question. Case studies (with the two exceptions already noted) rest upon an assumed synecdoche: The case should stand for a population. If this is not true, or if there is reason to doubt this assumption, then the utility of the case study is brought severely into question.

Fortunately, there is some safety in numbers. Insofar as case-study evidence is combined with cross-case evidence the issue of sample bias is mitigated. Indeed, the suspicion of case-study work that one finds in the social sciences today is, in my view, a product of a too-literal interpretation of the case-study method. A case study *tout* *(p. 677)* *court* is thought to mean a case study *tout seul*. Insofar as case studies and cross-case studies can be enlisted within the same investigation (either in the same study or by reference to other studies in the same subfield), problems of representativeness are less worrisome. This is the virtue of cross-level work, a.k.a. "triangulation."

11 Ambiguities

Before concluding, I wish to draw attention to two ambiguities in case-selection strategies in case-study research. The first concerns the admixture of several case- selection strategies. The second concerns the changing status of a case as a study proceeds.

Some case studies follow only one strategy of case selection. They are *typical*, *diverse*, *extreme*, *deviant*, *influential*, *crucial*, *pathway*, *most-similar*, or *most-different* research designs, as discussed. However, many case studies mix and match among these case-selection strategies. Indeed, insofar as all case studies seek representative samples, they are always in search of "typical" cases. Thus, it is common for writers to declare that their case is, for example, both extreme and typical; it has an extreme value on X_1 or Y but is not, in other respects, idiosyncratic. There is not much that one can say about these combinations of strategies except that, where the cases allow for a variety of empirical strategies, there is no reason not to pursue them. And where the same cases can serve several functions at once (without further effort on the researcher's part), there is little cost to a multi-pronged approach to case analysis.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

The second issue that deserves emphasis is the changing status of a case during the course of a researcher's investigation—which may last for years, if not decades. The problem is acute wherever a researcher begins in an exploratory mode and proceeds to hypothesis-testing (that is, she develops a specific X_1/Y proposition) or where the operative hypothesis or key control variable changes (a new causal factor is discovered or another outcome becomes the focus of analysis). Things change. And it is the mark of a good researcher to keep her mind open to new evidence and new insights. Too often, methodological discussions give the misleading impression that hypotheses are clear and remain fixed over the course of a study's development. Nothing could be further from the truth. The unofficial transcripts of academia— accessible in informal settings, where researchers let their guards down (particularly if inebriated)—are filled with stories about dead-ends, unexpected findings, and drastically revised theory chapters. It would be interesting, in this vein, to compare published work with dissertation prospectuses and fellowship applications. I doubt if the correlation between these two stages of research is particularly strong.

Research, after all, is about discovery, not simply the verification or falsification of static hypotheses. That said, it is also true that research on a particular topic should move from hypothesis generating to hypothesis-testing. This marks the progress of a (p. 678) field, and of a scholar's own work. As a rule, research that begins with an open-ended (X- or Y-centered) analysis should conclude with a determinate X_1/Y hypothesis.

The problem is that research strategies that are ideal for exploration are not always ideal for confirmation. The extreme-case method is inherently exploratory since there is no clear causal hypothesis; the researcher is concerned merely to explore variation on a single dimension (X or Y). Other methods can be employed in either an open- ended (exploratory) or a hypothesis-testing (confirmatory/disconfirmatory) mode. The difficulty is that once the researcher has arrived at a determinate hypothesis the originally chosen research design may no longer appear to be so well designed.

This is unfortunate, but inevitable. One cannot construct the perfect research design until (a) one has a specific hypothesis and (b) one is reasonably certain about what one is going to find “out there” in the empirical world. This is particularly true of observational research designs, but it also applies to many experimental research designs: Usually, there is a “good” (informative) finding, and a finding that is less insightful. In short, the perfect case-study research design is usually apparent only *ex post facto*.

There are three ways to handle this. One can explain, straightforwardly, that the initial research was undertaken in an exploratory fashion, and therefore not constructed to test the specific hypothesis that is—now—the primary argument. Alternatively, one can try to redesign the study after the new (or revised) hypothesis has been formulated. This may require additional field research or perhaps the integration of additional cases or variables that can be obtained through secondary sources or through consultation of experts. A final approach is to simply jettison, or de-emphasize, the portion of research

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

that no longer addresses the (revised) key hypothesis. A three-case study may become a two-case study, and so forth. Lost time and effort are the costs of this downsizing.

In the event, practical considerations will probably determine which of these three strategies, or combinations of strategies, is to be followed. (They are not mutually exclusive.) The point to remember is that revision of one's cross-case research design is *normal* and perhaps to be expected. Not all twists and turns on the meandering trail of truth can be anticipated.

12 Are There Other Methods of Case Selection?

At the outset of this chapter I summarized the task of case selection as a matter of achieving two objectives: representativeness (typicality) and variation (causal leverage). Evidently, there are other objectives as well. For example, one wishes to identify cases that are *independent* of each other. If chosen cases are affected by (p. 679) each other (sometimes known as Galton's problem or a problem of diffusion), this problem must be corrected before analysis can take place. I have neglected this issue because it is usually apparent to the researcher and, in any case, there are no simple techniques that might be utilized to correct for such biases. (For further discussion of this and other factors impinging upon case selection see Gerring 2001, 178–81.)

I have also disregarded *pragmatic/logistical* issues that might affect case selection. Evidently, case selection is often influenced by a researcher's familiarity with the language of a country, a personal entrée into that locale, special access to important data, or funding that covers one archive rather than another. Pragmatic considerations are often—and quite rightly—decisive in the case-selection process.

A final consideration concerns the *theoretical prominence* of a particular case within the literature on a subject. Researchers are sometimes obliged to study cases that have received extensive attention in previous studies. These are sometimes referred to as “paradigmatic” cases or “exemplars” (Flyvbjerg 2004, 427).

However, neither pragmatic/logistical utility nor theoretical prominence qualifies as a *methodological* factor in case selection. That is, these features of a case have no bearing on the validity of the findings stemming from a study. As such, it is appropriate to grant these issues a peripheral status in this chapter.

One final caveat must be issued. While it is traditional to distinguish among the tasks of case selection and case analysis, a close look at these processes shows them to be indistinct and overlapping. One cannot choose a case without considering the sort of analysis that it might be subjected to, and vice versa. Thus, the reader should consider choosing cases by employing the nine techniques laid out in this chapter *along with* any considerations that might be introduced by virtue of a case's quasi-experimental qualities, a topic taken up elsewhere (Gerring 2007, ch. 6).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

References

ABADIE, A., **DRUKKER**, D., **HERR**, J. L., and **IMBENS**, G. W. 2001. Implementing matching estimators for average treatment effects in Stata. *Stata Journal*, 1: 1–18.

ABBOTT, A. 2001. *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.

—— and **TSAY**, A. 2000. Sequence analysis and optimal matching methods in sociology. *Sociological Methods and Research*, 29: 3–33.

—— and **FORREST**, J. 1986. Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, 16: 471–94.

ACHEN, C. H., and **SNIDAL**, D. 1989. Rational deterrence theory and comparative case studies. *World Politics*, 41: 143–69.

ALLEN, W. S. 1965. *The Nazi Seizure of Power: The Experience of a Single German Town, 1930–1935*. New York: Watts.

ALMOND, G. A. 1956. Comparative political systems. *Journal of Politics*, 18: 391–409.

(p. 680) **AMENTA**, E. 1991. Making the most of a case study: theories of the welfare state and the American experience. Pp. 172–94 in *Issues and Alternatives in Comparative Social Research* ed. C. C. Ragin. Leiden: E. J. Brill.

BARRO, R. J. 1999. Determinants of democracy. *Journal of Political Economy*, 107: 158–83.

BELSEY, D. A., **KUH**, E., and **WELSCH**, R. E. 2004. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

BENNETT, A., **LEPGOLD**, J., and **UNGER**, D. 1994. Burden-sharing in the Persian Gulf War. *International Organization*, 48: 39–75.

BENTLEY, A. 1908/1967. *The Process of Government*. Cambridge, Mass.: Harvard University Press.

BRADY, H. E., and **COLLIER**, D. (eds.) 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Md.: Rowman and Littlefield.

BRAUMOELLER, B. F. 2003. Causal complexity and the study of politics. *Political Analysis*, 11: 209–33.

BREMAN, A., and **SHELTON**, C. 2001. Structural adjustment and health: a literature review of the debate, its role-players and presented empirical evidence. CMH Working Paper Series, Paper No. WG6: 6. WHO, Commission on Macroeconomics and Health.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

BRENNER, R. 1976. Agrarian class structure and economic development in pre-industrial Europe. *Past and Present*, 70: 30–75.

BROWNE, A. 1987. *When Battered Women Kill*. New York: Free Press.

BUCHBINDER, S., and **VITTINGHOFF**, E. 1999. HIV-infected long-term nonprogressors: epidemiology, mechanisms of delayed progression, and clinical and research implications. *Microbes Infect*, 1: 1113–20.

COHEN, M. R., and **NAGEL**, E. 1934. *An Introduction to Logic and Scientific Method*. New York: Harcourt, Brace and Company.

COLLIER, D., and **MAHONEY**, J. 1996. Insights and pitfalls: selection bias in qualitative research. *World Politics*, 49: 56–91.

COLLIER, R. B., and **COLLIER**, D. 1991/2002. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Notre Dame, Ind.: University of Notre Dame Press.

COLOMER, J. M. 1991. Transitions by agreement: modeling the Spanish way. *American Political Science Review*, 85: 1283–302.

CONVERSE, P. E., and **DUPEUX**, G. 1962. Politicization of the electorate in France and the United States. *Public Opinion Quarterly*, 16: 1–23.

COPPEDGE, M. J. 2004. The conditional impact of the economy on democracy in Latin America. Presented at the conference "Democratic Advancements and Setbacks: What Have We Learnt?", Uppsala University, June 11–13.

DE FELICE, E. G. 1986. Causal inference and comparative methods. *Comparative Political Studies*, 19: 415–37.

DESCH, M. C. 2002. Democracy and victory: why regime type hardly matters. *International Security*, 27: 5–47.

DEYO, F. (ed.) 1987. *The Political Economy of the New Asian Industrialism*. Ithaca, NY: Cornell University Press.

DION, D. 1998. Evidence and inference in the comparative case study. *Comparative Politics*, 30: 127–45.

ECKSTEIN, H. 1975. Case studies and theory in political science. In *Handbook of Political Science*, vii: *Political Science: Scope and Theory*, ed. F. I. Greenstein and N. W. Polsby. Reading, Mass.: Addison-Wesley.

(p. 681) **EGGAN**, F. 1954. Social anthropology and the method of controlled comparison. *American Anthropologist*, 56: 743–63.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

ELMAN, C. 2003. Lessons from Lakatos. In *Progress in International Relations Theory: Appraising the Field*, ed. C. Elman and M. F. Elman. Cambridge, Mass.: MIT Press.

—— 2005. Explanatory typologies in qualitative studies of international politics. *International Organization*, 59: 293–326.

EMIGH, R. 1997. The power of negative thinking: the use of negative case methodology in the development of sociological theory. *Theory and Society*, 26: 649–84.

EPSTEIN, L. D. 1964. A comparative study of Canadian parties. *American Political Science Review*, 58: 46–59.

ERTMAN, T. 1997. *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. Cambridge: Cambridge University Press.

ESPING-ANDERSEN, G. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.

FLYVBJERG, B. 2004. Five misunderstandings about case-study research. Pp. 420–34 in *Qualitative Research Practice*, ed. C. Seale, G. Gobo, J. F. Gubrium, and D. Silverman. London: Sage.

GEDDES, B. 1990. How the cases you choose affect the answers you get: selection bias in comparative politics. In *Political Analysis*, vol. ii, ed. J. A. Stimson. Ann Arbor: University of Michigan Press.

—— 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor: University of Michigan Press.

GEORGE, A. L., and **BENNETT**, A. 2005. *Case Studies and Theory Development*. Cambridge, Mass.: MIT Press.

—— and **SMOKE**, R. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. New York: Columbia University Press.

GERRING, J. 2001. *Social Science Methodology: A Criterial Framework*. Cambridge: Cambridge University Press.

—— 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

—— **THACKER**, S. and **MORENO**, C. 2005. Do neoliberal policies save lives? Unpublished manuscript.

GOERTZ, G. and **STARR**, H. (eds.) 2003. *Necessary Conditions: Theory, Methodology and Applications*. New York: Rowman and Littlefield.

—— and **LEVY**, J. (eds.) forthcoming. Causal explanations, necessary conditions, and case studies: World War I and the end of the Cold War. Manuscript.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

GOODIN, R. E. and **SMITSMAN**, A. 2000. Placing welfare states: the Netherlands as a crucial test case. *Journal of Comparative Policy Analysis*, 2: 39–64.

GUJARATI, D. N. 2003. *Basic Econometrics*, 4th edn. New York: McGraw-Hill.

HAMILTON, G. G. 1977. Chinese consumption of foreign commodities: a comparative perspective. *American Sociological Review*, 42: 877–91.

HAYNES, B. F. **PANTALEO**, G. and **FAUCI**, A. S. 1996. Toward an understanding of the correlates of protective immunity to HIV infection. *Science*, 271: 324–8.

HEMPEL, C. G. 1942. The function of general laws in history. *Journal of Philosophy*, 39: 35–48.

HO, D. E. **IMAI**, K. **KING**, G. and **STUART**, E. A. 2004. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Manuscript.

HOWARD, M. M. 2003. *The Weakness of Civil Society in Post-Communist Europe*. Cambridge: Cambridge University Press.

(p. 682) **HOWSON**, C. and **URBACH**, P. 1989. *Scientific Reasoning: The Bayesian Approach*. La Salle, Ill.: Open Court.

HUMPHREYS, M. 2005. Natural resources, conflict, and conflict resolution: uncovering the mechanisms. *Journal of Conflict Resolution*, 49: 508–37.

JENICEK, M. 2001. *Clinical Case Reporting in Evidence-Based Medicine*, 2nd edn. Oxford: Oxford University Press.

KARL, T. L. 1997. *The Paradox of Plenty: Oil Booms and Petro-states*. Berkeley: University of California Press.

KAZANCIGIL, A. 1994. The deviant case in comparative analysis: high stateness in comparative analysis. Pp. 213–38 in *Comparing Nations: Concepts, Strategies, Substance*, ed. M. Dogan and A. Kazancigil. Cambridge: Blackwell.

KEMP, K. A. 1986. Race, ethnicity, class and urban spatial conflict: Chicago as a crucial case *Urban Studies*, 23: 197–208.

KENDALL, P. L. and **WOLF**, K. M. 1949/1955. The analysis of deviant cases in communications research. In *Communications Research, 1948–1949*, ed. P. F. Lazarsfeld and F. N. Stanton. New York: Harper and Brothers. Reprinted as pp. 167–70 in *The Language of Social Research*, ed. P. F. Lazarsfeld and M. Rosenberg. New York: Free Press.

KENNEDY, C. H. 2005. *Single-case Designs for Educational Research*. Boston: Allyn and Bacon.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

KENNEDY, P. 2003. *A Guide to Econometrics*, 5th edn. Cambridge, Mass.: MIT Press.

KHONG, Y. F. 1992. *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, NJ: Princeton University Press.

KING, G. **KEOHANE**, R. O. and **VERBA**, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

LAKATOS, I. 1978. *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

LAZARSFELD, P. F. and **BARTON**, A. H. 1951. Qualitative measurement in the social sciences: classification, typologies, and indices. In *The Policy Sciences*, ed. D. Lerner and H. D. Lass- well. Stanford, Calif.: Stanford University Press.

LEVY, J. S. 2002. Qualitative methods in international relations. In *Evaluating Methodology in International Studies*, ed. F. P. Harvey and M. Brecher. Ann Arbor: University of Michigan Press.

LIJPHART, A. 1968. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. Berkeley: University of California Press.

—— 1969. Consociational democracy. *World Politics*, 21: 207–25.

—— 1971. Comparative politics and the comparative method. *American Political Science Review*, 65: 682–93.

—— 1975. The comparable cases strategy in comparative research. *Comparative Political Studies*, 8: 158–77.

LIPSET, S. M. 1959. Some social requisites of democracy: economic development and political development. *American Political Science Review*, 53: 69–105.

—— 1960/1963. *Political Man: The Social Bases of Politics*. Garden City, NY: Anchor.

—— 1968. *Agrarian Socialism: The Cooperative Commonwealth Federation in Saskatchewan. A Study in Political Sociology*. Garden City, NY: Doubleday.

—— **TROW**, M. A. and **COLEMAN**, J. S. 1956. *Union Democracy: The Internal Politics of the International Typographical Union*. New York: Free Press.

LYND, R. S. and **LYND**, H. M. 1929/1956. *Middletown: A Study in American Culture*. New York: Harcourt, Brace.

MAHONEY, J. and **GOERTZ**, G. 2004. The possibility principle: choosing negative cases in comparative research. *American Political Science Review*, 98: 653–69.

(p. 683) **MARTIN**, L. L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*.Princeton, NJ: Princeton University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

MAYO, D. G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

MECKSTROTH, T. 1975. "Most different systems" and "most similar systems:" a study in the logic of comparative inquiry. *Comparative Political Studies*, 8: 133–77.

MIGUEL, E. 2004. Tribe or nation: nation-building and public goods in Kenya versus Tanzania. *World Politics*, 56: 327–62.

MILL, J. S. 1843/1872. *The System of Logic*, 8th edn. London: Longmans, Green.

MONROE, K. R. 1996. *The Heart of Altruism: Perceptions of a Common Humanity*. Princeton, NJ: Princeton University Press.

MOORE, B., **JR**. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.

MORGAN, S. L. and **HARDING**, D. J. 2005. Matching estimators of causal effects: from stratification and weighting to practical data analysis routines. Manuscript.

MOULDER, F. V. 1977. *Japan, China and the Modern World Economy: Toward a Reinterpretation of East Asian Development ca. 1600 to ca. 1918*. Cambridge: Cambridge University Press.

MUNCK, G. L. 2004. Tools for qualitative research. Pp. 105–21 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Lanham, Md.: Rowman and Littlefield.

NJOLSTAD, O. 1990. Learning from history? Case studies and the limits to theory-building. Pp. 220–46 in *Arms Races: Technological and Political Dynamics*, ed. O. Njolstad. Thousand Oaks, Calif.: Sage.

PATTON, M. Q. 2002. *Qualitative Evaluation and Research Methods*. Newbury Park, Calif.: Sage.

POPPER, K. 1934/1968. *The Logic of Scientific Discovery*. New York: Harper and Row.

—— 1963. *Conjectures and Refutations*. London: Routledge and Kegan Paul.

POSNER, D. 2004. The political salience of cultural difference: why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review*, 98: 529–46.

PRZEWORSKI, A. and **TEUNE**, H. 1970. *The Logic of Comparative Social Inquiry*. New York: John Wiley.

QUEEN, S. 1928. Round table on the case study in sociological research. *Publications of the American Sociological Society, Papers and Proceedings*, 22: 225–7.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

RAGIN, C. C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.

—— 2004. Turning the tables. Pp. 123–38 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Lanham, Md.: Rowman and Littlefield.

REILLY, B. 2000–1. Democracy, ethnic fragmentation, and internal conflict: confused theories, faulty data, and the "crucial case" of Papua New Guinea. *International Security*, 25: 162–85.

—— and **PHILLPOT**, R. 2003. "Making democracy work" in Papua New Guinea: social capital and provincial development in an ethnically fragmented society. *Asian Survey*, 42: 906–27.

ROGOWSKI, R. 1995. The role of theory and anomaly in social-scientific inference. *American Political Science Review*, 89: 467–70.

ROHLFING, I. 2004. Have you chosen the right case? Uncertainty in case selection for single case studies. Working Paper, International University, Bremen.

ROSENBAUM, P. R. 2004. Matching in observational studies. In *Applied Bayesian Modeling and Causal Inference from an Incomplete-data Perspective*, ed. A. Gelman and X.-L. Meng. New York: John Wiley.

—— and **SILBER**, J. H. 2001. Matching and thick description in an observational study of mortality after surgery. *Biostatistics*, 2: 217–32.

ROSS, M. 2001. Does oil hinder democracy? *World Politics*, 53: 325–61.

(p. 684) **SAGAN**, S. D. 1995. *Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press.

SEKHON, J. S. 2004. Quality meets quantity: case studies, conditional probability and counter- factuals. *Perspectives in Politics*, 2: 281–93.

SHAFER, M. D. 1988. *Deadly Paradigms: The Failure of U.S. Counterinsurgency Policy*. Princeton, NJ: Princeton University Press.

SKOCPOL, T. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.

—— and **SOMERS**, M. 1980. The uses of comparative history in macrosocial inquiry. *Comparative Studies in Society and History*, 22: 147–97.

STINCHCOMBE, A. L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace.

SWANK, D. H. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States*. Cambridge: Cambridge University Press.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

TENDLER, J. 1997. *Good Government in the Tropics*. Baltimore: Johns Hopkins University Press.

TRUMAN, D. B. 1951. *The Governmental Process*. New York: Alfred A. Knopf.

TSAI, L. 2007. *Accountability without Democracy: How Solidary Groups Provide Public Goods in Rural China*. Cambridge: Cambridge University Press.

VAN EVERA, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.

WAHLKE, J. C. 1979. Pre-behavioralism in political science.*American Political Science Review*, 73: 9–31.

YASHAR, D. J. 2005. *Contesting Citizenship in Latin America: The Rise of Indigenous Movements and the Postliberal Challenge*. Cambridge: Cambridge University Press.

YIN, R. K. 2004. *Case Study Anthology*. Thousand Oaks, Calif.: Sage.

Notes:

(1) Gujarati (2003); Kennedy (2003). Interestingly, the potential of cross-case statistics in helping to choose cases for in-depth analysis is recognized in some of the earliest discussions of the case-study method (e.g. Queen 1928, 226).

(2) This expands on Mill (1843/1872, 253), who wrote of scientific enquiry as twofold: “either inquiries into the cause of a given effect or into the effects or properties of a given cause.”

(3) This method has not received much attention on the part of qualitative methodologists; hence, the absence of a generally recognized name. It bears some resemblance to J. S. Mill's Joint Method of Agreement and Difference (Mill 1843/1872), which is to say a mixture of most-similar and most-different analysis, as discussed below. Patton (2002, 234) employs the concept of “maximum variation (heterogeneity) sampling.”

(4) More precisely, George and Smoke (1974, 534, 522–36, ch. 18; see also discussion in Collier and Mahoney 1996, 78) set out to investigate causal pathways and discovered, through the course of their investigation of many cases, these three causal types. Yet, for our purposes what is important is that the final sample includes at least one representative of each “type.”

(5) For further examples see Collier and Mahoney (1996); Geddes (1990); Tendler (1997).

(6) Traditionally, methodologists have conceptualized cases as having “positive” or “negative” values (e.g. Emigh 1997; Mahoney and Goertz 2004; Ragin 2000, 60; 2004, 126).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

(7) Geddes (1990); King, Keohane, and Verba (1994). See also discussion in Brady and Collier (2004); Collier and Mahoney (1996); Rogowski (1995).

(8) The exception would be a circumstance in which the researcher intends to disprove a deterministic argument (Dion 1998).

(9) Geddes (2003, 131). For other examples of casework from the annals of medicine see "Clinical reports" in the *Lancet*, "Case studies" in *Canadian Medical Association Journal*, and various issues of the *Journal of Obstetrics and Gynecology*, often devoted to clinical cases (discussed in Jenicek 2001, 7). For examples from the subfield of comparative politics see Kazancigil (1994).

(10) For a discussion of the important role of anomalies in the development of scientific theorizing see Elman (2003); Lakatos (1978). For examples of deviant-case research designs in the social sciences see Amenta (1991); Coppedge (2004); Eckstein (1975); Emigh (1997); Kendall and Wolf (1949/1955).

(11) For examples of the crucial-case method see Bennett, Lepgold, and Unger (1994); Desch (2002); Goodin and Smitsman (2000); Kemp (1986); Reilly and Phillpot (2003). For general discussion see George and Bennett (2005); Levy (2002); Stinchcombe (1968, 24–8).

(12) A third position, which purports to be neither Popperian or Bayesian, has been articulated by Mayo (1996, ch. 6). From this perspective, the same idea is articulated as a matter of "severe tests."

(13) It should be noted that Tsai's conclusions do not rest solely on this crucial case. Indeed, she employs a broad range of methodological tools, encompassing case-study and cross-case methods.

(14) See also the discussion in Eckstein (1975) and Lijphart (1969). For additional examples of case studies disconfirming general propositions of a deterministic nature see Allen (1965); Lipset, Trow, and Coleman (1956); Njolstad (1990); Reilly (2000–1); and discussion in Dion (1998); Rogowski (1995).

(15) Granted, insofar as case-study analysis provides a window into causal mechanisms, and causal mechanisms are integral to a given theory, a single case may be enlisted to confirm or disconfirm a proposition. However, if the case study upholds a posited pattern of *X/Y* covariation, and finds fault only with the stipulated causal mechanism, it would be more accurate to say that the study forces the *reformulation* of a given theory, rather than its confirmation or disconfirmation. See further discussion in the following section.

(16) Sometimes, the most-similar method is known as the "method of difference," after its inventor (Mill 1843/1872). For later treatments see Cohen and Nagel (1934); Eggan (1954); Gerring (2001, ch. 9); Lijphart (1971; 1975); Meckstroth (1975); Przeworski and Teune (1970); Skocpol and Somers (1980).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

(17) For good introductions see Ho et al. (2004); Morgan and Harding (2005); Rosenbaum (2004); Rosenbaum and Silber (2001). For a discussion of matching procedures in Stata see Abadie et al. (2001).

(18) The most-different method is also sometimes referred to as the "method of agreement," following its inventor, J. S. Mill (1843/1872). See also De Felice (1986); Gerring (2001, 212–14); Lijphart (1971; 1975); Meckstroth (1975); Przeworski and Teune (1970); Skocpol and Somers (1980). For examples of this method see Collier and Collier (1991/2002); Converse and Dupeux (1962); Karl (1997); Moore (1966); Skocpol (1979); Yashar (2005, 23). However, most of these studies are described as *combining* most-similar and most-different methods.

(19) In the following discussion I treat the terms social capital, civil society, and civic engagement interchangeably.

(20) E.g. Collier and Collier (1991/2002); Karl (1997); Moore (1966); Skocpol (1979); Yashar (2005, 23). Karl (1997), which affects to be a most-different system analysis (20), is a particularly clear example of this. Her study, focused ostensibly on petro-states (states with large oil reserves), makes two sorts of inferences. The first concerns the (usually) obstructive role of oil in political and economic development. The second sort of inference concerns variation *within* the population of petro-states, showing that some countries (e.g. Norway, Indonesia) manage to avoid the pathologies brought on elsewhere by oil resources. When attempting to explain the constraining role of oil on petro-states, Karl usually relies on contrasts between petro-states and nonpetro-states (e.g. ch. 10). Only when attempting to explain differences among petro-states does she restrict her sample to petro-states. In my opinion, very little use is made of the most-different research design.

(21) This was recognized, at least implicitly, by Mill (1843/1872, 258–9). Skepticism has been echoed by methodologists in the intervening years (e.g. Cohen and Nagel 1934, 251–6; Gerring 2001; Skocpol and Somers 1980). Indeed, explicit defenses of the most-different method are rare (but see De Felice 1986).

(22) Another way of stating this is to say that X is a "nontrivial necessary condition" of Y.

(23) Wahlke (1979, 13) writes of the failings of the "behavioralist" mode of political science analysis: "It rarely aims at generalization; research efforts have been confined essentially to case studies of single political systems, most of them dealing …with the American system."

John Gerring

John Gerring is Professor of Political Science, Boston University.

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: Goldsmiths College, University of London; date: 07 March 2019